AN ELASTIC TRAFFIC POLICER FOR MPEG VIDEO STREAMS

Emre Gündüzhan

Nortel
20410 Century Blvd.
Germantown, MD 20874

ABSTRACT

Multimedia applications like MPEG-2 video produce packets with multiple levels of importance from application quality point of view. We propose a traffic policing mechanism for such applications, called elastic policing, which provides more flexibility to more important packets. We describe a token bucket based implementation of elastic policing using additional "loan buckets", and verify its performance by simulation studies. We observe significant improvement in objective video quality without increasing traffic admitted into the network in the long term.

1. INTRODUCTION

Emerging services such as multimedia over IP and virtual private networks demand careful resource allocation in an IP network. It is necessary to know bounds on offered traffic to be able to allocate sufficient resources for these services. The traffic profile of a service level agreement (SLA) between a customer and a service provider specifies these bounds, usually in the form of several token bucket (TB) parameters. The in-profile traffic, which conforms to the bounds, must be serviced by the network according to the desired QoS specified in the SLA, while any excess out-of-profile traffic is forwarded without any guarantees.

In a QoS-enabled IP network incoming traffic is first policed to make sure that it conforms to the traffic profile. The policing mechanism works at network edge, and marks each packet with a color. All congestion control mechanisms inside the network use this color to determine which packets should be dropped during congestion. The policing mechanism is usually some variant of a TB algorithm [2]. A three-color TB marks in-profile traffic as green, some out-of-profile traffic as yellow and some as red according to a second boundary also specified in the SLA. Yellow traffic is forwarded with a higher drop-precedence than green traffic. Red traffic is either forwarded with a higher drop-precedence than yellow traffic, or it is discarded immediately.

The existing policing mechanisms do not consider the importance of individual packets. In many applications, different types of packets carry different levels of information from a service quality point of view. One example is video traffic. It is known that losing packets from certain types of video frames have more adverse effect on video quality than losing packets from other types of frames [4, 9]. More specifically for MPEG-2 video, discarding an I-packet causes larger quality degradation than discarding a P-packet, which in turn causes larger quality degradation than discarding a B-packet[1]. This is because other frames depend on I- and P-frames, causing error propagation, but no other frames depend on a B-frame. Another example is layered encoding of voice or video, where dropping a packet from a base layer results in a loss in all layers, while dropping a packet at a higher layer can partially be mitigated by using lower layers. In all these cases, existing mechanisms police according to rate only, so packets at all importance levels get the same drop-precedence characteristics.

In the rest of this paper we focus on MPEG-2 video, although the proposed mechanisms can be used with other applications as well. When VBR video traffic is policed by a regular TB algorithm, it is possible that all three types of packets can occasionally be out-of-profile and therefore all three types of packets can be discarded irrespective of their type. Alternatively one can again police the traffic by a regular TB algorithm but does not downgrade or discard important packets (i.e. I and/or P types). However, in this case the average rate of admitted traffic into the network will be larger than specified in the SLA and may exceed the amount that can be handled. Therefore a new policing mechanism is needed which maintains conformance to the traffic profile on average, but also has some flexibility for important packets to reduce the possibility of these packets marked as yellow or red. We call such a mechanism *elastic policing* because of the additional flexibility. We propose a new TB-based implementation for elastic policing in this paper and investigate its performance.

There is a plethora of TB algorithms in the literature. Streams with multiple levels of priority were considered in [5] in the context of a TB traffic shaper for ATM networks.

[1] We use the following terminology: MPEG-2 operates on *frames*, each of which is carried by multiple *packets* in the network.

0-7803-9332-5/05/$20.00

This shaper uses different number of tokens for different priority levels, and employs a jump-ahead algorithm to prevent blocking of high priority cells. However using a similar algorithm for policing a video stream requires apriori knowledge of the rate of each frame type, which is not practical. Furthermore packet reordering by the jump-ahead can be undesirable.

Undesired effects of the strict nature of standard TB policing algorithms are also well-known, and many variations were proposed to introduce some flexibility. In [8] and [3] two algorithms were proposed to distribute unused bandwidth[2] in a group of connections among the connections that require more bandwidth. In this way each individual connection can exceed its allocation to some extent but the group allocation is strictly enforced. In [1] a method called credit banking was proposed to allow a stream to borrow bandwidth from its future. Although the idea of borrowing is similar to the elastic policing idea that we propose, in credit banking the borrowed bandwidth is only used to smooth the policy decisions rather than providing a differentiation among packets.

In elastic policing we mark traffic streams in such a way that the marked stream conforms to the rates imposed by an SLA on average, and at the same time important packets within the stream get better colors than less important packets. We propose a practical implementation of elastic policing based on a TB algorithm with additional loan buckets. Loan buckets allow important packets to borrow bandwidth, which is paid back by less important packets. This mechanism increases the burst size of in-profile traffic, but maintains the same average rate.

In the rest of this document, we first summarize TB policers in Section 2. Then we describe TB with loan in Section 3 and show our simulation results in Section 4. We conclude the paper in Section 5.

2. TOKEN BUCKET POLICERS

Internet standard [2] defines a two-rate policer. It can operate in either color-aware or color-blind mode, depending on whether incoming traffic is already marked or not. We focus on color-blind mode in this document, but the proposed algorithms can also be used in color-aware mode with simple modifications. We also explain the algorithms for the more complex two-rate case. It is straightforward to apply them to the simpler single-rate case.

The two-rate policer in [2] is based on four parameters: committed information rate (CIR), committed burst size (CBS), peak information rate (PIR), and peak burst size (PBS). There are two token buckets, C and P, with number of tokens at any time given by N_C and N_P, respectively. The sizes of both buckets are limited by the allowable burst sizes, so that $N_C \leq CBS$ and $N_P \leq PBS$. Both buckets are initially full and receive tokens at regular intervals with rates CIR and PIR, respectively. When a packet of size S arrives, it is marked as follows:

- if $N_P - S < 0$, then the packet is red and neither N_C nor N_P is decremented,
- else if $N_C - S < 0$, then the packet is yellow and N_P is decremented by S,
- else, the packet is green and both N_C and N_P are decremented by S.

Let t_0 be any instant when both buckets are full. Then, for any time $t \geq t_0$, green traffic is limited by

$$B_G = (t - t_0) * CIR + CBS, \qquad (1)$$

and the aggregate of green and yellow traffic is limited by

$$B_{GY} = (t - t_0) * PIR + PBS. \qquad (2)$$

3. TOKEN BUCKET WITH LOAN

The basic policing operation described in the previous section imposes a shaping constraint on the traffic stream, but it does not distinguish between individual packets within the traffic stream. On the other hand, there can be different levels of importance among the packets of the same stream, such as I, P, and B frame packets of MPEG-2 video. The elastic policing mechanism we propose provides some flexibility for important packets. One way of policing a stream elastically is to use additional loan buckets with a regular TB algorithm. The main idea is that when an important packet is to be marked as yellow or red, it can borrow bandwidth and marked as green instead. Later, a less important packet pays the debt and marked as yellow or red instead of the earlier important packet.

We assume three levels of importance in traffic streams, with level 1 most important and level 3 least important. For instance, the levels 1, 2, and 3 may correspond to I, P, and B packets, respectively, for MPEG-2 video. We have regular token buckets C and P with all the parameters as described in the previous section. These buckets are initially full. We also have three loan buckets, L_{gy}, L_{gr}, and L_{yr}. Each bucket L_{mn} counts the amount of loan borrowed for color m from color n. All loan buckets are initially empty. The number of tokens in a loan bucket L_{mn} is given by M_{mn}. Each loan bucket has two thresholds, $T_{mn,1}$ and $T_{mn,2}$, with $T_{mn,2} <= T_{mn,1}$. We borrow bandwidth for packets at importance level i ($i = 1, 2$) as long as the debt remains less than $T_{mn,i}$. Level-3 packets are used to pay the debt. The thresholds, $T_{mn,i}$, can be seen as a measure of credibility, i.e. as the debt increases it gets more difficult to borrow additional bandwidth. The token buckets C and P are incremented regularly, at rates CIR and PIR, up to

[2] We use "bandwidth" synonymously with "bit rate" in this paper.

the maxima CBS and PBS. When a packet of size S arrives, it is marked according to its importance level, i, as follows.

a) $i = 1$ or $i = 2$:

if $N_P - S < 0$ then

- if $M_{gr} + S < T_{gr,i}$, then the packet is green and M_{gr} is incremented by S,
- else if $M_{yr} + S < T_{yr,i}$, then the packet is yellow and M_{yr} is incremented by S,
- else, the packet is red;

else if $N_C - S < 0$ then

- if $M_{gy} + S < T_{gy,i}$, then the packet is green, M_{gy} is incremented by S and N_P is decremented by S,
- else, the packet is yellow and N_P is decremented by S;

else, the packet is green and both N_P and N_C are decremented by S.

b) $i = 3$:

if $N_P - S < 0$ then the packet is red;

else if $N_C - S < 0$ then

- if $M_{yr} > 0$, then the packet is red and both N_P and M_{yr} are decremented by S,
- else, the packet is yellow and N_P is decremented by S;

else

- if $M_{gr} > 0$, then the packet is red and M_{gr}, N_P, and N_C are decremented by S,
- else if $M_{gy} > 0$, then the packet is yellow and M_{gy}, N_P, and N_C are decremented by S,
- else, the packet is green and both N_P and N_C are decremented by S.

In summary, when a level-1 or level-2 packet arrives and it is about to be marked different than green, the relevant credit thresholds are checked. If there is enough credit, then the packet is marked as green, or yellow instead of red, and the relevant loan counter is increased. When a level-3 packet arrives and it is about to be marked different than red, relevant loan counters are checked. If there is any loan previously borrowed, then the packet is colored with the color loan was borrowed from, and the counter is decreased.

Compared to the admitted traffic shapes in (1) and (2), the additional instantaneous green burst allowed by our algorithm is limited by $T_{gy,1} + T_{gr,1}$, and the additional instantaneous green + yellow burst allowed by our algorithm is limited by $T_{yr,1} + T_{gr,1}$. However, these are short-term bursts, which are eliminated as long as there are sufficient level-3 packets in the stream. Therefore on average the admitted traffic does not exceed the original limits (1) and (2).

4. SIMULATION RESULTS

We tested the performance of elastic policing in two simulations. In the first simulation we used four video models based on [6] and [7] derived from four different short video clips. Two clips were 25 fps and two clips were 29.97 fps. The GOP structures were either 12 or 15 frames long with I to P distance of 3. The average bit rate of each model was 4 Mb/s. Ten independent instances of each model were multiplexed, resulting in 40 source streams, and transmitted over a 1 Gb/s link.

	Regular TB			TB with Loan		
Type	Green	Yellow	Red	Green	Yellow	Red
I	89.81	10.07	0.12	99.99	0.01	0
P	89.64	10.23	0.12	99.62	0.38	0
B	89.89	9.98	0.13	72.87	26.79	0.34

Table 1. Marking percentages of different packet types.

	Reg. TB Rate (Mb/s)			TB /w Ln. Rate (Mb/s)		
Color	Avg	St Dv	Max	Avg	St Dv	Max
Green	163.3	26.4	231.6	163.3	28.5	263.2
Gr+Ye	182.2	37.1	339.9	182.2	37.1	339.9

Table 2. Admittance rate of green and green+yellow traffic averaged over 1-ms intervals.

The aggregate video stream was policed with a regular TB with parameters $CIR = 200$ Mb/s, $CBS = 4500$ bytes, $PIR = 400$ Mb/s, and $PBS = 6000$ bytes. The same stream was also policed by a TB with loan with the same TB parameters. The level-1 and level-2 thresholds for all loan buckets were set to 4500 bytes and 2250 bytes, respectively. The frame type information (I, P, B) was mapped to importance levels (1, 2, 3).

The resulting marking percentages are shown in Table 1. Regular TB marks all three packet types with the same characteristics, as expected. TB with loan reduces the percentage of yellow I-packets and P-packets almost to zero, and eliminates red I-packets and P-packets completely.

We also measured the rates of the green traffic and the aggregate of green and yellow traffic using 1-ms intervals, and then calculated the average, maximum, and the standard deviation of the measured rates. The results are shown in Table 2. TB with loan produces the same average rate for both green and green+yellow traffic. The standard deviation and maximum rate increase slightly due to the additional flexibility provided for I-packets and P-packets. This increase was not observed for the green+yellow traffic, because the amount of red traffic was too small to begin with (less than 0.13%) so the additional allowed traffic was not significant.

In our second simulation we investigated shaped VBR traffic. We assumed that each source stream consisted of a number of scenes with three different content material. Scene duration was exponentially distributed. The content of a new scene was random and equally likely. The scene content models corresponded to a fast motion (football) clip at 9 Mb/s, a slow motion clip with high texture at 7 Mb/s, and a slow motion clip with low texture at 3 Mb/s. 60 independent source streams were multiplexed with average rate 380 Mb/s.

The aggregate video stream was first shaped at 400 Mb/s with a varying shaper buffer size. When the shaper buffer was full, it was emptied at 800 Mb/s until the buffer size reduces below some threshold. The resulting traffic was policed either by regular TB or by TB with loan. The TB parameters were $CIR = 400$ Mb/s, $PIR = 500$ Mb/s, $CBS = PBS = 4500$ bytes.

We plotted the marking ratio vs. shaper buffer size for the fast motion scenes in Figure 1. Once again regular TB does not distinguish between frame types, while TB with loan improves I- and P- frame marking at the expense of B-frames. There were no red I- or P- frames with TB with loan. We observed similar results for the other two scenes.

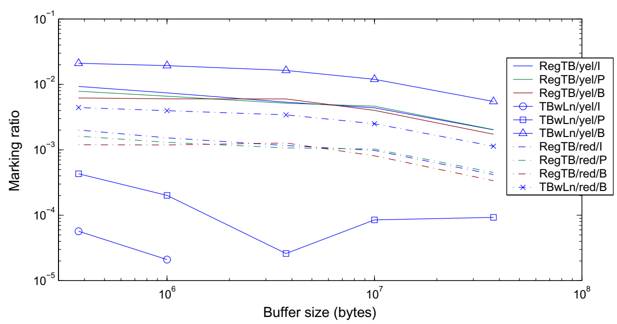

Fig. 1. Improvement in I- and P-packet marking by TB with loan.

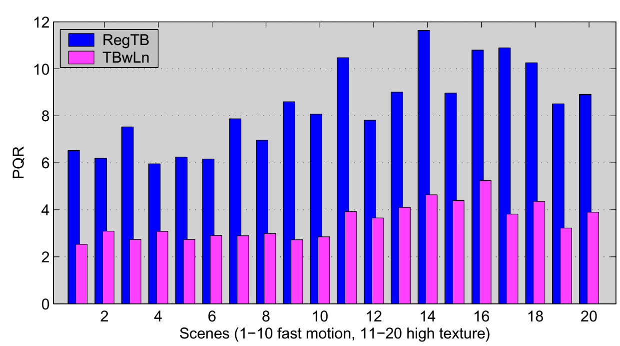

Fig. 2. Improvement in objective quality by TB with loan.

We also investigated the impact of elastic policing on perceived quality of video. For this purpose we used a commercial objective video quality measurement tool, Tektronix PQA300. We randomly picked ten fast motion scenes and ten high texture scenes. We assumed red packets were dropped in the network, while green and yellow packets passed, and measured the quality of resulting video clips against the original scenes. Figure 2 shows the result in terms of picture quality rating (PQR) score. PQR increases with increased distortion. PQR=0 corresponds to the original image, PQR=1 corresponds to small perceptual impact, and PQR$\geq$3 corresponds to almost always observable distortion. The figure shows that TB with loan improves the PQR between 3 and 7, which is significant.

5. CONCLUDING REMARKS

We proposed a new policing mechanism, called elastic policing, which can be used for streams with multiple importance levels such as video traffic. We also proposed a new TB-based implementation for elastic policing, with additional loan buckets.

Simulation results for video traffic showed that our method improves the drop precedence of important packets substantially, which results in better quality video. The improvement in perceived video quality is significant during the interval that degradation occurs. Simulation results show an improvement of 3 to 7 in PQR score. Our method does not change the amount of total in-profile traffic admitted in the long term, but increases the instantaneous amount. The allowed instantaneous increase is adjustable, by modifying the loan bucket thresholds.

6. REFERENCES

[1] Y. Gong and I. F. Akyıldız, “Average rate enforcement for real-time traffic in ATM networks,” *Technical Report, Georgia Institute of Technology*, Feb. 1992.

[2] J. Heinanen and R. Guerin, “A two rate three color marker,” *RFC 2698*, http://www.ietf.org/rfc/rfc2698.txt, Sept. 1999.

[3] J. S. M. Ho, H. Uzunalioğlu and I. F. Akyıldız, “Cooperating leaky bucket for average rate enforcement of VBR video traffic in ATM networks,” *Proc. of IEEE INFOCOM’95*, April 1995.

[4] M. R. Ito and Y. Bai, “A packet discard scheme for loss control in IP networks with MPEG video traffic,” *8th Int. Conf. on Comm. Systems, ICCS 2002*, 25-28 Nov. 2002.

[5] C. P. Ko et al, “Using token allocations in a leaky bucket scheme,” *Military Communications Conference, MILCOM94*, 25 Oct. 1994.

[6] M. M. Krunz and A. M. Makowski, “Modeling video traffic using $M/G/\infty$ input processes: a compromise between Markovian and LRD models,” *IEEE J. on Sel. Areas in Comm.*, Vol. 16, No. 5, June 1998.

[7] “A simple model for MPEG video traffic,” *ICME 2000*, Vol. 1, 30 July-2 Aug. 2000.

[8] S.-L. Wu and W.-S. E. Chen, “The token-bank leaky bucket mechanism for group connections in ATM networks,” *Proc. of Int. Conf. on Network Protocols*, 29 Oct. 1 Nov. 1996.

[9] H. Zhao, N. Ansari and Y. Q. Shi, “Transmission of real-time video over IP differentiated services,” *Electronics Letters*, Vol. 38, Sep. 2002.